

PARTIAL TRANSLATION OF JP 11(1999)-306354 A

Publication Date: November 5, 1999

Title of the Invention: METHOD AND DEVICE FOR PICTURE  
RECOGNITION

Patent Application Number: 10-110454

Filing Date: April 21, 1998

Inventors: Megumi YAMAOKA et al.

Applicant: MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.

#### Claims

[Claim 1] A method for picture recognition, comprising: dividing an input picture into local regions; extracting a local region resembling each input local region from a learning picture database in which learning pictures are previously registered; totalizing the number of matching between relative coordinates of input local regions in the input picture and those of corresponding learning local regions in learning pictures with respect to each learning picture; and selecting the learning picture having the highest totalized value.

[Claim 2] A device for picture recognition, comprising: picture division means for dividing an input picture into local regions; resembling window extraction means for extracting a local region resembling each input local region from a learning picture database in which learning pictures are previously registered; totalizing means for totalizing the number of matching between relative coordinates of input local regions in the input picture and those of corresponding learning local regions in learning pictures with respect to each learning picture; and object determination means for selecting the learning picture having the highest totalized value.

[Claim 3] A device for picture recognition for recognizing an object substance in an input picture by searching for a picture resembling the input picture

from previously obtained high-volume learning picture data, comprising: picture division means for dividing an input picture into local regions; a learning picture database in which pictures of various substances are divided into windows with the same size as a division size of the picture division means, and stored as learning windows together with their positional information on the pictures; resembling window extraction means for extracting learning windows resembling the respective input windows divided by the picture division means, and outputting paired input windows and learning windows together on the basis of a learning picture to which the learning window belongs; learning means for prescribing any three windows belonging to the same learning picture stored in the learning picture database as a base, and storing a combination of coordinates calculated from a positional relationship on the learning picture and the base into a learning window coordinate database on a learning picture basis; voting means for extracting all the combinations of three windows to be a base with respect to learning windows belonging to the same learning picture input from the resembling window extraction means, obtaining coordinates in a space formed by the input base and a base of windows belonging to the same learning picture from the base and a positional relationship of windows on an input picture, and totalizing the number of matching between the calculated coordinates and the coordinates in the learning window database of the same base/window; and selecting the learning picture having the highest totalized value.

[Claim 4] A device for picture recognition according to claim 3, wherein the learning means includes a similar kind of picture information database in which, in a case where there are a plurality of learning pictures of the similar kind of substance, each learning window of the learning pictures is previously stored by being associated with a learning window of one representative learning picture, and based on the similar kind of picture information in the similar kind of picture information database, the voting means converts a learning window belonging to the learning picture of the

similar kind of substance input from the resembling window extraction means into a learning window of one representative learning picture and extracting all the combinations of three windows to be a base with respect to learning windows belonging to the input similar kind of picture.

[Claim 5] A device for picture recognition for recognizing an object substance in an input picture by searching for a picture resembling the input picture from previously obtained high-volume learning picture data, comprising: picture division means for dividing an input picture into local regions; resembling window extraction means for extracting learning windows resembling the respective input windows divided by the picture division means, and outputting paired input windows and learning windows together on the basis of a learning picture to which the learning window belongs; learning means having a learning picture database in which pictures of various substances are divided into windows with the same size as a division size of the picture division part and stored as learning windows together with their positional information on a picture, and simultaneously, any one point is prescribed as an attention point representing the position of a substance in a picture on a learning picture basis and its positional information is stored; a window selection part for extracting all the combinations of two windows with respect to learning windows belonging to the same learning picture input from the resembling window extraction means; voting means for, on the basis of a combination of windows extracted in the window selection part, estimating a position coordinate of the attention point on the input picture from position coordinates of two input windows and a positional relationship of the same window/attention point input from the learning picture database on the learning picture, obtaining a vector formed by two windows from the position coordinates on the input picture with respect to each extracted combination of windows, calculating an angle formed by the vector and a vector obtained from the position coordinate of the same window on the learning picture stored in the learning picture database, and prescribing the input coordinate value and the vector

angle input from the vector angle calculation part as one combination and totalizing the number of combinations of windows with the same value on a learning picture basis; and object determination means for selecting the learning picture having the highest totalized value.

[Claim 6] A device for picture recognition according to claim 5, wherein the learning part includes a similar kind of picture information database in which, in a case where there are a plurality of learning pictures of the similar kind of substance, each learning window of the learning pictures is stored by being associated with a learning window of one representative learning picture, and the voting means includes a similar kind of picture mixing part for, based on the similar kind of picture information in the similar kind of picture information database, converting a learning window belonging to a learning picture of the similar kind of substance input from the resembling window extraction part into a learning window belonging to one representative learning picture and outputting it, and extracts all the combinations of two windows with respect to learning windows belonging to the similar kind of picture input from the similar kind of picture mixing part.

(Page 3, right column, lines 13-26)

[Embodiments of the Invention] According to the invention defined by claim 1, an input picture is divided into local regions, a local region resembling each input local region is extracted from a learning picture database in which learning pictures are previously registered, the number of matching between relative coordinates of input local regions in the input picture and those of corresponding learning local regions in learning pictures are totalized with respect to each learning picture, and the learning picture having the highest totalized value is selected. Because of this, the present invention has a function that relative coordinates of local regions in the input picture are compared with those of learning local regions in learning pictures resembling the input picture local regions, a learning picture having matched relative coordinates is considered as a substance of the input

picture, and a parallel movement amount and a rotation angle of a substance in the input picture are calculated from the difference in absolute coordinates of corresponding local regions between the learning picture and the input picture.

(Page 4, right column, lines 27-46)

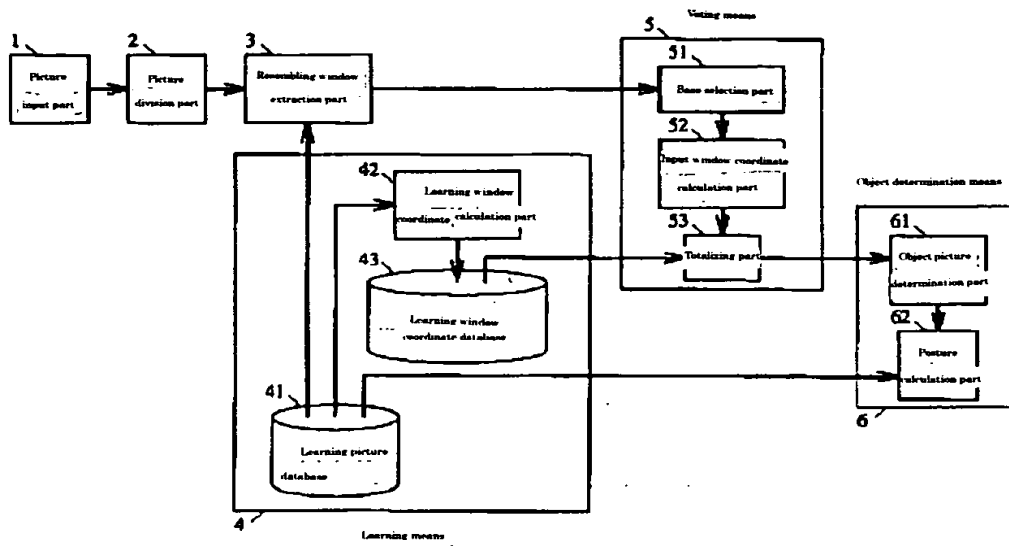
[0013] According to the invention defined by claim 6, in a device of picture recognition of claim 5, the learning part includes a similar kind of picture information database in which, in a case where there are a plurality of learning pictures of the similar kind of substance, each learning window of the learning pictures is stored by being associated with a learning window of one representative learning picture, and the voting means includes a similar kind of picture mixing part for, based on the similar kind of picture information in the similar kind of picture information database, converting a learning window belonging to a learning picture of the similar kind of substance input from the resembling window extraction part into a learning window belonging to one representative learning picture and outputting it, and extracts all the combinations of two windows with respect to learning windows belonging to the similar kind of picture input from the similar kind of picture mixing part. Because of this, the present invention has a function that, in a case where there are a plurality of similar kind of learning pictures, they are allowed to be represented by one learning picture; thereafter, an angle is calculated that is formed by the vector of two local regions in the input picture and the vector created by two learning picture local regions belonging to the same learning picture resembling the local regions of the input picture, and a learning picture is estimated as a substance of the input picture, in which the angle of two local regions is constant, and the position of the substance in the input picture estimated by the learning local region resembling each input local region is most constant.

(Page 7, right column, lines 11-41)

[0034] When window picture data is input from a picture division part 2, a

resembling window extraction part 3 calculates the difference (e.g., sum of the square of the difference of the respective pixel values) between the input window picture data and all the learning window picture data in a learning picture database 41, and extracts the learning data with the smallest difference. In the learning picture database 41, as shown in Figure 5, pictures of various substances are previously divided into windows with the same size as that of an input window picture, and stored together with a window number and a position coordinate of the center of a window. When extracting the respective learning windows most resembling the input windows from the learning picture database 41, the resembling window extraction part 3 summarizes them on the basis of a learning picture to which learning windows belong, and, as shown in Figure 6, outputs them as a pair of a learning window identification number and a center coordinate of the corresponding input window (1103). Herein, the learning window identification number consists of a learning picture number to which a window belongs and a number representing the position in the learning picture, as shown in Figure 5. Furthermore, by being associated with coordinates of the input window, each learning window can be considered as a vector. In the similar kind of picture information database 44, learning windows containing the same part of each substance are stored by being associated with each other as similar kind of picture information, with respect to a learning picture of the same kind of substance stored in the learning picture database 41. For example, a learning picture 3 shown in Figure 5 and a learning picture 29 shown in Figure 10 are learning pictures of a sedan in the lateral direction. The learning window 3 of the learning picture 3 and the learning window 27 of the learning picture 29 are considered as pictures containing rear part windows of both the sedans, so that they are considered as corresponding learning windows. Similarly, as shown in Figure 11, the similar kind of picture information database 44 stores the correspondence in learning window between the learning picture 3 and the learning picture 29, and information that a representative picture of the similar kind of learning pictures 3, 29, 62, ... is the learning picture 3.

Fig. 1



(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11306354 A**

(43) Date of publication of application: **05 . 11 . 99**

(51) Int. Cl.

**G06T 7/00**  
**G06T 1/00**

(21) Application number: **10110454**

(22) Date of filing: **21 . 04 . 98**

(71) Applicant: **MATSUSHITA ELECTRIC IND CO LTD**

(72) Inventor: **YAMAOKA MEGUMI**  
**NAGAO KENJI**

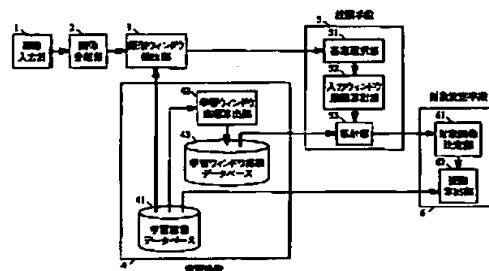
(54) **METHOD AND DEVICE FOR PICTURE RECOGNITION**

COPYRIGHT: (C)1999,JPO

(57) Abstract:

**PROBLEM TO BE SOLVED:** To provide a device which correctly recognizes an object even in the case that the object in an input picture is moving or rotating at the time of recognizing the object in the input picture on a learning picture most resembling the input picture.

**SOLUTION:** This device is provided with a picture division means 2 which divides a picture inputted from a picture input part into local regions, a resembling window extraction means 3 which extracts local resembling each input local region from a learning picture database, a totalizing means 5 which totalizes the number of matching between relative coordinates of input local regions in the input picture and those of corresponding learning local in learning pictures with respect to each learning picture, and an object determination means 6 which selects the learning picture having the highest totalized value, thus performing picture recognition with a high reliability even in the occurrence of changes in position and attitude of the object in the input picture.





(11)特許出願公開番号

特開平11-306354

(43)公開日 平成11年(1999)11月5日

(51) IntCl.<sup>5</sup>

**識別記号**

F I

G O B T 7/00  
1/00

G O 6 F 15/70  
15/62

4 6 5 A  
3 8 0

審査請求 未請求 請求項の数7 OL (全 16 頁)

(21)出願番号 特願平10-110454

(22)出願日 平成10年(1998)4月21日

(71)出願人 000005821

松下電器産業株式会社  
大阪府門真市大字門真1006番地

(72)発明者 山岡 めぐみ

神奈川県川崎市多摩区東三田3丁目10番1号 松下技研株式会社内

(72)発明者 長尾 健司

神奈川県川崎市多摩区東三田3丁目10番1号 松下技研株式会社内

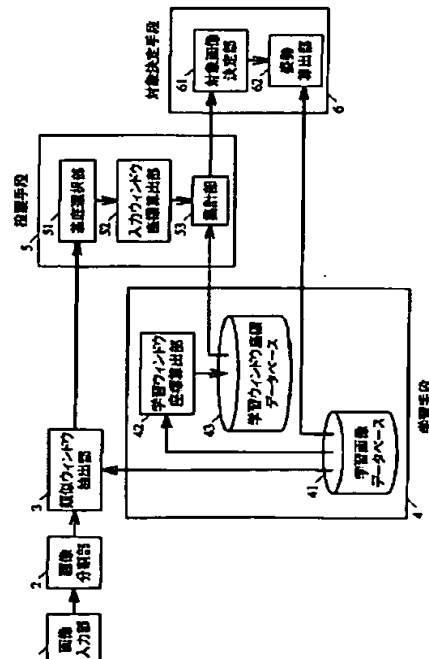
(74)代理人 弁理士 滝本 智之 (外1名)

(54) 【発明の名称】 画像認識方法及び画像認識装置

(57) 【要約】

【課題】 入力画像と最も類似する学習画像で入力画像中の物体を認識する際に、入力画像中の物体が移動していたり回転している場合にも正しく物体を認識する装置を提供することを目的とする。

【解決手段】 画像入力部1から入力した画像を局所領域に分割する画像分割手段2と、各入力局所領域に対して学習画像データベースから類似する局所領域を抽出する類似ウィンドウ抽出手段3と、入力局所領域同士の入力画像における相対座標と対応する学習局所領域同士の学習画像における相対座標が一致するものの数を学習画像ごとに集計する集計手段5と、集計値の最も高い学習画像を選択する対象決定手段6を備えることにより、入力画像中の物体の位置・姿勢が変動している場合にも信頼性の高い画像認識を行うことができる。



1

## 【特許請求の範囲】

【請求項1】 入力した画像を局所領域に分割し、各入力局所領域に対して予め学習画像を登録した学習画像データベースから類似する局所領域を抽出して、前記入力局所領域同士の入力画像における相対座標と対応する学習局所領域同士の学習画像における相対座標が一致するものの数を学習画像ごとに集計値を求め、前記集計値の最も高い学習画像を選択することを特徴とする画像認識方法。

【請求項2】 入力した画像を局所領域に分割する画像分割手段と、各入力局所領域に対して、予め学習画像を登録した学習画像データベースから類似する局所領域を抽出する類似ウィンドウ抽出手段と、前記入力局所領域同士の入力画像における相対座標と対応する学習局所領域同士の学習画像における相対座標が一致するものの数を学習画像ごとに集計する集計手段と、前記集計値の最も高い学習画像を選択する対象決定手段とを備えることを特徴とする画像認識装置。

【請求項3】 予め採取しておいた大量の学習画像データの中から、入力画像に近い画像を探索して、入力画像中の対象物体を認識する装置において、入力した画像を局所領域に分割する画像分割手段と、様々な物体の画像を前記画像分割手段での分割サイズと同じサイズのウィンドウに分割して画像上での位置情報と共に学習ウィンドウとして格納している学習画像データベースと、前記画像分割手段で分割したそれぞれの入力ウィンドウに対して類似した学習ウィンドウを抽出して入力ウィンドウと学習ウィンドウの対を学習ウィンドウが属する学習画像ごとにまとめて出力する類似ウィンドウ抽出手段と、前記学習画像データベースに格納されている同じ学習画像に属するウィンドウについて任意の3ウィンドウを基底とし学習画像上での位置関係から算出した座標と基底の組を学習画像ごとに学習ウィンドウ座標データベースへ格納する学習手段と、前記類似ウィンドウ抽出手段から入力した同じ学習画像に属する学習ウィンドウについて基底となる3ウィンドウの組み合わせを全て抽出し、入力した基底と同じ学習画像に属するウィンドウの基底が張る空間での座標を基底とウィンドウの入力画像上での位置関係から求め、算出した座標と同じ基底・ウィンドウによる前記学習ウィンドウデータベースでの座標とを比べて一致するものの数を集計する投票手段と、前記集計値の最も高い学習画像を選択する対象決定手段を備えたことを特徴とする画像認識装置。

【請求項4】 学習手段は、予め同種の物体の学習画像が複数ある場合にそれらの各学習ウィンドウをその中の代表的な1つの学習画像中の学習ウィンドウと対応付けて格納している同種画像情報データベースを有し、投票手段は前記同種画像情報データベースの同種画像情報に基いて前記類似ウィンドウ抽出手段から入力した同種の物体の学習画像に属する学習ウィンドウを代表的な一つ

2

の学習画像の学習ウィンドウに変換し、入力した同種の画像に属する学習ウィンドウについて基底となる3ウィンドウの組み合わせを全て抽出することを特徴とする請求項3記載の画像認識装置。

【請求項5】 予め採取しておいた大量の学習画像データの中から、入力画像に近い画像を探索して、入力画像中の対象物体を認識する装置において、入力した画像を局所領域に分割する画像分割手段と、前記画像分割手段で分割したそれぞれの入力ウィンドウに対して類似した学習ウィンドウを抽出して入力ウィンドウと学習ウィンドウの対を学習ウィンドウが属する学習画像ごとにまとめて出力する類似ウィンドウ抽出手段と、様々な物体の画像を前記画像分割部での分割サイズと同じサイズのウィンドウに分割して画像上での位置情報と共に学習ウィンドウとして格納し同時に学習画像ごとに任意の一点を画像中の物体の位置を示す注目点としてその位置情報を格納している学習画像データベースを持つ学習手段と、前記類似ウィンドウ抽出手段から入力した同じ学習画像に属する学習ウィンドウについて2ウィンドウの組み合わせを全て抽出するウィンドウ選択部と、前記ウィンドウ選択部で抽出したウィンドウの組ごとに注目点の入力画像上での位置座標を2入力ウィンドウの位置座標と前記学習画像データベースから入力した同じウィンドウ・注目点の学習画像上での位置関係から推定し、抽出した各ウィンドウの組について入力画像上での位置座標から2ウィンドウが成すベクトルを求め、前記学習画像データベースに格納されている同じウィンドウの学習画像上での位置座標から求めたベクトルとのなす角を算出し、入力した座標値と前記ベクトル角算出部から入力したベクトル角を1つの組にして同じ値を持つウィンドウの組の数を学習画像ごとに集計する投票手段と、前記集計値の最も高い学習画像を選択する対象決定手段とを備えたことを特徴とする画像認識装置。

【請求項6】 学習部は、同種の物体の学習画像が複数ある場合にそれらの各学習ウィンドウをその中の代表的な一つの学習画像中の学習ウィンドウと対応付けて格納している同種画像情報データベースを有し、投票手段は前記同種画像情報データベースの同種画像情報に基いて前記類似ウィンドウ抽出部から入力した同種の物体の学習画像に属する学習ウィンドウを代表的な一つの学習画像に属する学習ウィンドウに変換して出力する同種画像混合部と、前記同種画像混合部から入力した同種の画像に属する学習ウィンドウについて2ウィンドウの組み合わせを全て抽出することを特徴とする請求項5記載の画像認識装置。

【請求項7】 コンピュータにより画像認識を行うプログラムであって、入力した画像を局所領域に分割し、各入力局所領域に対して予め学習画像を登録した学習画像データベースから類似する局所領域を抽出して、前記入力局所領域同士の入力画像における相対座標と対応する

学習局所領域同士の学習画像における相対座標が一致するものの数を学習画像ごとに集計値を求め、前記集計値の最も高い学習画像を選択する画像認識プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、入力画像が、予め作成済の学習画像データベース中のどの画像と近いかを判定することにより、入力画像上に表示されている物体が何であるかを認識する画像認識方法及び画像認識装置並びに画像認識プログラムを記録した記録媒体に関するものである。

【0002】

【従来の技術】従来の画像認識装置は、特開平 9-21610号公報に記載されたものが知られている。

【0003】図 16 は、従来の画像認識装置のブロック構成図を示しており、画像を入力する画像入力部 11 と、抽出対象物の局所モデルを予め格納しているモデル記憶部 12 と、入力画像の各部分画像について各局所モデルとのマッチングを行うマッチング処理部 13 と、入力画像の各部分画像がどの程度局所モデルに一致しているかによって画像の位置情報も含めたパラメータ空間で抽出対象物の位置を確率的に表示し統合する局所情報統合部 14 と、パラメータ空間内で最も確立の高い部分を抽出して入力画像内での抽出対象物の位置を判別して出力する物体位置決定部 15 から構成されている。

【0004】

【発明が解決しようとする課題】このような従来の画像認識装置は、抽出対象物が入力画像と学習画像の間で画像面内平行移動している場合には適用できるが、平行移動と画像面内回転移動が同時に生じている場合は認識できないという課題を有していた。

【0005】本発明は、上記従来の課題を解決するもので、入力画像の対象物体がモデルと比べて画像面内で回転している場合にも対象物体を認識し、その位置とモデルに対する回転角を推定することを目的とする。

【0006】

【課題を解決するための手段】この課題を解決するために本発明は、第 1 に画像を入力する画像入力部と、前記画像入力部から入力した画像をウィンドウに分割する画像分割部と、前記画像分割部で分割したそれぞれの入力ウィンドウに対して類似した学習ウィンドウを抽出する類似ウィンドウ抽出部と、前記類似ウィンドウ抽出部から入力した同じ学習画像に属する学習ウィンドウについて基底となる 3 ウィンドウを選択した後入力画像上での位置関係から基底以外のウィンドウを基底が張る空間座標で求める入力ウィンドウ座標算出部と、前記入力ウィンドウ座標算出部で算出した座標と同じ基底・ウィンドウの座標を学習画像上での位置関係から算出したものを前記入力ウィンドウ座標算出部で算出した座標と比較し

て一致するものの数を学習画像ごとに集計する集計部を備えたものである。

【0007】これにより、まず、入力ウィンドウと類似する学習ウィンドウを抽出し、次に、同じ学習画像に属する学習ウィンドウの学習画像上での相対座標と、対応する入力ウィンドウの入力画像上での相対座標を比較して、一致するものの数を学習画像ごとに集計することによって、入力画像に最も一致する学習画像を決定し、更に、その学習画像に属する学習ウィンドウと対応する入力ウィンドウの絶対座標の差から入力画像中の物体の回転角と入力画像上での位置を推定することができる。

【0008】

【発明の実施の形態】本発明の請求項 1 に記載の発明は、入力した画像を局所領域に分割し、各入力局所領域に対して予め学習画像を登録した学習画像データベースから類似する局所領域を抽出して、前記入力局所領域同士の入力画像における相対座標と対応する学習局所領域同士の学習画像における相対座標が一致するものの数を学習画像ごとに集計値を求め、前記集計値の最も高い学習画像を選択するもので、入力画像中の各局所領域の相対座標と、入力画像局所領域と類似する学習画像局所領域の学習画像上の相対座標を比較して、相対座標が一致する学習画像を入力画像の物体であるとし、学習画像と入力画像の対応する局所領域の絶対座標の違いから、入力画像中の物体の平行移動量と回転角を算出するという作用を有する。

【0009】請求項 2 に記載の発明は、入力した画像を局所領域に分割する画像分割手段と、各入力局所領域に対して、予め学習画像を登録した学習画像データベースから類似する局所領域を抽出する類似ウィンドウ抽出手段と、前記入力局所領域同士の入力画像における相対座標と対応する学習局所領域同士の学習画像における相対座標が一致するものの数を学習画像ごとに集計する集計手段と、前記集計値の最も高い学習画像を選択する対象決定手段と備えたもので、入力画像中の各局所領域の相対座標と、入力画像局所領域と類似する学習画像局所領域の学習画像上の相対座標を比較して、相対座標が一致する学習画像を入力画像の物体であるとし、学習画像と入力画像の対応する局所領域の絶対座標の違いから、入力画像中の物体の平行移動量と回転角を算出するという作用を有する。

【0010】請求項 3 に記載の発明は、予め採取しておいた大量の学習画像データの中から、入力画像に近い画像を探索して、入力画像中の対象物体を認識する装置において、入力した画像を局所領域に分割する画像分割手段と、様々な物体の画像を前記画像分割手段での分割サイズと同じサイズのウィンドウに分割して画像上での位置情報と共に学習ウィンドウとして格納している学習画像データベースと、前記画像分割手段で分割したそれぞれの入力ウィンドウに対して類似した学習ウィンドウを抽

出して入力ウィンドウと学習ウィンドウの対を学習ウィンドウが属する学習画像ごとにまとめて出力する類似ウィンドウ抽出手段と、前記学習画像データベースに格納されている同じ学習画像に属するウィンドウについて任意の3ウィンドウを基底とし学習画像上での位置関係から算出した座標と基底の組を学習画像ごとに学習ウィンドウ座標データベースへ格納する学習手段と、前記類似ウィンドウ抽出手段から入力した同じ学習画像に属する学習ウィンドウについて基底となる3ウィンドウの組み合わせを全て抽出し、入力した基底と同じ学習画像に属するウィンドウの基底が張る空間での座標を基底とウィンドウの入力画像上での位置関係から求め、算出した座標と同じ基底・ウィンドウによる前記学習ウィンドウデータベースでの座標とを比べて一致するものの数を集計する投票手段と、前記集計値の最も高い学習画像を選択する対象決定手段を備えたもので、入力画像中の各局所領域の相対座標と、入力画像局所領域と類似する学習画像局所領域の学習画像上の相対座標を比較して、相対座標が一致する学習画像を入力画像の物体であるとし、学習画像と入力画像の対応する局所領域の絶対座標の違いから、入力画像中の物体の平行移動量と回転角を算出するという作用を有する。

【0011】請求項4に記載の発明は、請求項3記載の画像認識装置において、学習手段は、予め同種の物体の学習画像が複数ある場合にそれらの各学習ウィンドウをその中の代表的な1つの学習画像中の学習ウィンドウと対応付けて格納している同種画像情報データベースを有し、投票手段は前記同種画像情報データベースの同種画像情報に基いて前記類似ウィンドウ抽出手段から入力した同種の物体の学習画像に属する学習ウィンドウを代表的な一つの学習画像の学習ウィンドウに変換し、入力した同種の画像に属する学習ウィンドウについて基底となる3ウィンドウの組み合わせを全て抽出するもので、同種の学習画像が複数ある場合に、それを一つの学習画像で代表させた後入力画像中の各局所領域の相対座標と、入力画像局所領域と類似する学習画像局所領域の学習画像上の相対座標を比較して、相対座標が一致する学習画像を入力画像の物体の種類を表す画像であるとし、学習画像と入力画像の対応する局所領域の絶対座標の違いから、入力画像中の物体の平行移動量と回転角を算出するという作用を有する。

【0012】請求項5に記載の発明は、予め採取しておいた大量の学習画像データの中から、入力画像に近い画像を探索して、入力画像中の対象物体を認識する装置において、入力した画像を局所領域に分割する画像分割手段と、前記画像分割手段で分割したそれぞれの入力ウィンドウに対して類似した学習ウィンドウを抽出して入力ウィンドウと学習ウィンドウの対を学習ウィンドウが属する学習画像ごとにまとめて出力する類似ウィンドウ抽出手段と、様々な物体の画像を前記画像分割部での分割

サイズと同じサイズのウィンドウに分割して画像上での位置情報と共に学習ウィンドウとして格納し同時に学習画像ごとに任意の一点を画像中の物体の位置を示す注目点としてその位置情報を格納している学習画像データベースを持つ学習手段と、前記類似ウィンドウ抽出手段から入力した同じ学習画像に属する学習ウィンドウについて2ウィンドウの組み合わせを全て抽出するウィンドウ選択部と、前記ウィンドウ選択部で抽出したウィンドウの組ごとに注目点の入力画像上での位置座標を2入力ウィンドウの位置座標と前記学習画像データベースから入力した同じウィンドウ・注目点の学習画像上での位置関係から推定し、抽出した各ウィンドウの組について入力画像上での位置座標から2ウィンドウが成すベクトルを求め、前記学習画像データベースに格納されている同じウィンドウの学習画像上での位置座標から求めたベクトルのなす角を算出し、入力した座標値と前記ベクトル角算出部から入力したベクトル角を1つの組にして同じ値を持つウィンドウの組の数を学習画像ごとに集計する投票手段と、前記集計値の最も高い学習画像を選択する対象決定手段とを備えたもので、入力画像中の各2局所領域が作るベクトルと、その局所領域と類似する同じ学習画像に属する2学習画像局所領域が作るベクトルのなす角を算出して、各2局所領域のなす角が一定であり、かつ、各入力局所領域と類似した学習局所領域が推定する入力画像中の物体の位置が最も一定となる学習画像を入力画像の物体であると推定するという作用を有する。

【0013】請求項6に記載の発明は、請求項5記載の画像認識装置において、学習部は、同種の物体の学習画像が複数ある場合にそれらの各学習ウィンドウをその中の代表的な一つの学習画像中の学習ウィンドウと対応付けて格納している同種画像情報データベースを有し、投票手段は前記同種画像情報データベースの同種画像情報に基いて前記類似ウィンドウ抽出部から入力した同種の物体の学習画像に属する学習ウィンドウを代表的な一つの学習画像に属する学習ウィンドウに変換して出力する同種画像混合部と、前記同種画像混合部から入力した同種の画像に属する学習ウィンドウについて2ウィンドウの組み合わせを全て抽出するもので、同種の学習画像が複数ある場合に、それを一つの学習画像で代表させた後、入力画像中の各2局所領域が作るベクトルと、その局所領域と類似する同じ学習画像に属する2学習画像局所領域が作るベクトルのなす角を算出して、各2局所領域のなす角が一定であり、かつ、各入力局所領域と類似した学習局所領域が推定する入力画像中の物体の位置が最も一定となる学習画像を入力画像の物体であると推定するという作用を有する。

【0014】請求項7に記載の発明は、コンピュータにより画像認識を行うプログラムであって、入力した画像を局所領域に分割し、各入力局所領域に対して予め学習画像を登録した学習画像データベースから類似する局所

領域を抽出して、前記入力局所領域同士の入力画像における相対座標と対応する学習局所領域同士の学習画像における相対座標が一致するものの数を学習画像ごとに集計値を求め、前記集計値の最も高い学習画像を選択する画像認識プログラムを記録した記録媒体から、コンピュータに読み込み実行することにより、入力画像中の各局所領域の相対座標と、入力画像局所領域と類似する学習画像局所領域の学習画像上の相対座標を比較して、相対座標が一致する学習画像を入力画像の物体であるとし、学習画像と入力画像の対応する局所領域の絶対座標の違いから、入力画像中の物体の平行移動量と回転角を算出するという作用を有する。

【0015】以下、本発明の実施の形態について、図1から図14を用いて説明する。

（実施の形態1）図1は、本発明の実施の形態1における画像認識装置のブロック構成図を示している。図1において、1は認識したい対象物の画像データを入力する画像入力部、2は画像入力部1で入力した画像を局所ウィンドウに分割して出力する画像分割部、3は画像分割部2で分割した各入力ウィンドウ画像に対して類似する学習ウィンドウをデータベースから抽出して学習ウィンドウが属する画像ごとにまとめて、対応する入力ウィンドウと共に出力する類似ウィンドウ抽出部、4は認識したい物体のモデルを予め作成しておく学習手段、41は種々の物体のモデル画像である学習画像を、画像分割部2で作成する局所ウィンドウと同じサイズの局所ウィンドウに分割して学習ウィンドウとして格納している学習画像データベース、42は学習画像データベースに格納されている各学習画像ごとに、任意の3つの学習ウィンドウを抽出して、その3学習ウィンドウの位置座標を基底としたときのそれ以外の学習ウィンドウの相対座標を求める学習ウィンドウ座標算出部、43は学習ウィンドウ座標算出部42で求めた相対座標データを格納している学習ウィンドウ座標データベース、5は類似ウィンドウ抽出部3で抽出した学習ウィンドウの、学習画像上での相対位置と、対応する入力ウィンドウの入力画像上での相対位置を比較して、一致するものの数を集計する投票手段、51は類似ウィンドウ抽出部3が出力する同一学習画像に属する学習ウィンドウの中から、基底となる3つの学習ウィンドウの組み合わせを全て抽出して、対応する入力ウィンドウと共に基底学習ウィンドウ・基底入力ウィンドウとして出力する基底選択部、52は基底選択部51から入力した各基底について、同じ学習画像に属する他の学習ウィンドウの相対座標を、対応する入力ウィンドウの対応する基底入力ウィンドウに対する入力画像上での相対座標で表して、座標値と、学習ウィンドウ・基底学習ウィンドウと、対応する入力ウィンドウ・基底入力ウィンドウを出力する入力ウィンドウ座標算出部、53は入力ウィンドウ座標算出部52から入力した学習ウィンドウ・基底学習ウィンドウに対して学習ウ

ィンドウ座標データベース43に格納されている相対座標と入力ウィンドウ座標算出部52から入力した座標とを比較して、一致するものの数を基底ごとに集計して基底と集計値を出力する集計部、6は投票手段5の集計結果を受けて入力画像中の対象物とその姿勢を決定する対象決定手段、61は集計部53から入力した集計値のうち最大の値を持つ基底を出力し、その基底が属する学習画像中の物体を入力画像の物体であるとみなす対象画像決定部、62は対象画像決定部61が選択した基底について、基底学習ウィンドウの学習画像上での位置と、基底入力ウィンドウの入力画像上での位置とを比較して、入力画像中の対象の位置と回転角を算出する姿勢算出部である。

【0016】また、図2はコンピュータにより画像認識装置を実現した場合のブロック構成図であり、201はコンピュータ、202はCPU、203はメモリ、204はキーボード及びディスプレイ、205は画像認識プログラムを読み込むためのFDユニット、206～208はI/Fユニット、209はCPUバス、210は画像を取り込むためのカメラ、211は予め蓄積されている画像を取り込むための画像データベース、212は種々の物体のモデル画像である学習画像を局所ウィンドウに分割して学習ウィンドウとして格納している学習画像データベース、213は相対座標データを格納している学習ウィンドウ座標データベース、215は得られた物体の平行移動量と回転角はI/Fユニットを介して出力する出力端子で構成されている。

【0017】以上のように構成された画像認識装置について、以下その動作を図3のフローチャートを用いて説明する。図4は、入力画像の一例、図5は、学習画像の一例、図6は、類似ウィンドウ抽出部3が出力するデータの一例、図7は、集計部53が出力する集計結果の一例である。

【0018】認識対象となる画像データを画像入力部1（カメラ210または画像データベース211）から入力する（601）。画像分割部2は、図4に示すように、その画像から一定サイズのウィンドウを任意画素移動させて順次抽出し、ウィンドウ画像データをウィンドウの中心点の座標とともに出力する（602）。

【0019】学習画像データベース41（学習画像データベース212）には、予め、種々の物体の画像が、図5に示すように、入力ウィンドウ画像と同じサイズのウィンドウに区切られ、ウィンドウ番号とウィンドウの中心点の位置座標とともに格納されている。

【0020】類似ウィンドウ抽出部3は、画像分割部2からウィンドウ画像データを入力すると、学習画像データベース41（学習画像データベース212）の全ての学習ウィンドウ画像データと差、例えば、各画素値の差の二乗の和、を算出して、最も差の小さいものを抽出する。類似ウィンドウ抽出部3は、全ての入力ウィンドウ

に対してそれぞれ最も類似した学習ウィンドウを学習画像データベース41から抽出すると、学習ウィンドウが属する学習画像ごとにまとめて、図6に示すように、学習ウィンドウ識別番号と、対応する入力ウィンドウの中心座標の対で出力する(603)。ここで、学習ウィンドウ識別番号は、図5に示すように、属する学習画像番号と学習画像中の位置を示す番号から成り、また、入力ウィンドウの座標と対応付けることにより、各学習ウィンドウは、ベクトルとみなすことができる。

【0021】基底選択部51は、学習ウィンドウ識別番号と座標を入力すると、まず、同じ学習画像に属する学習ウィンドウを抽出する(604)。次に、抽出した学習ウィンドウから基底とする3つのウィンドウの組み合わせを全て作る(605)。例えば、図7の画像3に属する基底は、(3-3, 3-4, 3-11), (3-3, 3-4, 3-12), ...となる。

【0022】次に、各基底ウィンドウの組に対して、基底ウィンドウが張る空間での基底以外のウィンドウの座標を求める(606)。例えば、図7の基底(3-3, 3-4, 3-11)の基底ベクトルは、3-3を原点とすると、3-3, 3-4から成るベクトル(119-40, 86-100)=(79, -14)(ベクトルe1)と3-3, 3-11から成るベクトル(54-40, 179-100)=(14, 79)(ベクトルe2)である。よって、基底以外のウィンドウ3-5(198, 72)の3-3を原点としたときの座標(198-40, 72-100)=(158, -28)をベクトルe1とベクトルe2で表すと、(2\*ベクトルe1) + (0\*ベクトルe2)となり、3-5の相対座標は、(2, 0)となる。

【0023】ここで、学習ウィンドウ座標算出部42は、予め、全ての学習画像の全ての基底に対する各学習ウィンドウの学習画像上での相対座標を、各ウィンドウの中心点の座標から算出して、学習ウィンドウ座標データベース43(学習ウィンドウ座標データベース213)に格納している。例えば、図6における3-3と3-4と3-5と3-11の学習画像上での中心座標は、図5に示したように、それぞれ(200, 40)(280, 40)(360, 40)(200, 120)であるので、3-3, 3-4, 3-11を基底としたときの基\*

$$e_i = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad \dots (1)$$

(ただし、e<sub>i</sub>は、入力画像の基底ベクトル、

e<sub>1</sub>は、学習画像上の基底ベクトルを示す。)

【0029】なお、得られた物体の平行移動量と回転角は、1/Fユニットを介して出力端子215から出力される。

【0030】(実施の形態2)図8は、本発明の実施の形態2における画像認識装置のブロック構成図を示す。図8において、1は認識したい対象物の画像データを入力する画像入力部、2は画像入力部1で入力した画像を

\* 基底ベクトルは、3-3を原点とすると(80, 0)=ベクトルe1, (0, 80)=ベクトルe2となる。よって、3-5:(360, 40)の、3-3を原点としたときの座標は(160, 0)となり、相対座標は、(2\*ベクトルe1) + (0\*ベクトルe2)すなわち、(2, 0)となる。

【0024】集計部53は、(606)で算出した座標(座標A)と同じ基底・同じウィンドウによる座標を学習ウィンドウ座標データベース43から抽出して(座標B)、座標Aが座標Bの近傍にあれば(607)、その基底の得点として1点加算する(608)。例えば、上記の、基底(3-3, 3-4, 3-11)・ウィンドウ3-5の場合は、座標Aと座標Bが一致するので、基底(3-3, 3-4, 3-11)の得点に1点加算する。

【0025】図7に示すように、全ての画像の全ての基底について(604)から(608)までの処理が終了したら、その学習画像の全ての学習ウィンドウについて処理が終了したかを判定する(609)。次に、その学習画像の全ての基底について処理が終了したかを判定する(610)。次に、全ての学習画像について処理が終了したかを判定する(611)。

【0026】対象画像決定部6は、学習画像ごとに得点を合計して最も高得点の学習画像の物体を入力画像の対象物体とする(612)。

【0027】更に、姿勢算出部62は、その学習画像に属する高得点の基底の入力画像上の座標と、学習画像データベース41に格納されている学習画像上の座標から、物体の平行移動量と回転角を求める(613)。例えば、図7の高得点の基底(3-3, 3-4, 3-11)の場合、3-3の学習画像上の座標(200, 40)(図5)と、入力画像上の座標(40, 100)(図4)から、平行移動量を(-160, 60)と求め、回転角θは、3-4の学習画像上の座標(280, 40)(図9)と入力画像上の座標(119, 86)(図10)から、(数1)を解いて、θ=10度と求めることができる。

【0028】

【数1】

局所ウィンドウに分割して出力する画像分割部、3は画像分割部2で分割した各入力ウィンドウ画像に対して類似する学習ウィンドウをデータベースから抽出して学習ウィンドウが属する画像ごとにまとめて、対応する入力ウィンドウと共に出力する類似ウィンドウ抽出部、4は認識したい物体のモデルを予め作成しておく学習手段、41は種々の物体のモデル画像である学習画像を、画像

## 11

分割部2で作成する局所ウィンドウと同じサイズの局所ウィンドウに分割して学習ウィンドウとして格納している学習画像データベース、42は学習画像データベースに格納されている各学習画像ごとに、任意の3つの学習ウィンドウを抽出して、その3学習ウィンドウの位置座標を基底としたときのそれ以外の学習ウィンドウの相対座標を求める学習ウィンドウ座標算出部、43は学習ウィンドウ座標算出部42で求めた相対座標データを格納している学習ウィンドウ座標データベース、44は学習画像データベース41中の同じ種類の複数の学習画像についてそれらに属する各学習ウィンドウを代表的な一つの学習画像の同じ部分の学習ウィンドウと対応づけて格納している同種画像情報データベース、5は類似ウィンドウ抽出部3で抽出した学習ウィンドウの、学習画像上での相対位置と、対応する入力ウィンドウの入力画像上での相対位置を比較して、一致するものの数を集計する投票手段、50は同種画像情報データベース44に格納されている同種画像情報に基づいて類似ウィンドウ抽出部3が出力する同一学習画像に属する学習ウィンドウを同じ種類の代表的な1学習画像に含まれる学習ウィンドウに変換して出力する同種画像混合部、51は同種画像混合部50が出力する同一種類の学習画像に属する学習ウィンドウの中から、基底となる3つの学習ウィンドウの組み合わせを全て抽出して、対応する入力ウィンドウと共に基底学習ウィンドウ・基底入力ウィンドウとして出力する基底選択部、52は基底選択部51から入力した各基底について、同じ学習画像に属する他の学習ウィンドウの相対座標を、対応する入力ウィンドウの対応する基底入力ウィンドウに対する入力画像上での相対座標で表して、座標値と、学習ウィンドウ・基底学習ウィンドウと、対応する入力ウィンドウ・基底入力ウィンドウを出力する入力ウィンドウ座標算出部、53は入力ウィンドウ座標算出部52から入力した学習ウィンドウ・基底学習ウィンドウに対して学習ウィンドウ座標データベース43に格納されている相対座標と入力ウィンドウ座標算出部52から入力した座標とを比較して、一致するものの数を基底ごとに集計して基底と集計値を出力する集計部、6は投票手段5の集計結果を受けて入力画像中の対象物とその姿勢を決定する対象決定手段、61は集計部53から入力した集計値のうち最大の値を持つ基底を出力し、その基底が属する学習画像中の物体を入力画像の物体であるとみなす対象画像決定部、62は対象画像決定部61が選択した基底について、基底学習ウィンドウの学習画像上での位置と、基底入力ウィンドウの入力画像上での位置とを比較して、入力画像中の対象の位置と回転角を算出する姿勢算出部である。

【0031】以上のように構成された画像認識装置について、以下その動作を図9に示すフローチャートを用いて説明する。

【0032】図4は入力画像の一例、図5は学習画像の

## 12

一例、図6は類似ウィンドウ抽出部3が出力するデータの一例、図7は集計部53が出力する集計結果の一例、図10は学習画像データベース41に格納されている同種画像の一例図、11は同種画像情報データベース44に格納されている同種画像情報の一例である。

【0033】認識対象となる画像データが画像入力部1から入力する(1101)。画像分割部2は、図4に示すように、その画像から一定サイズのウィンドウを順次抽出して、ウィンドウ画像データをウィンドウの中心点の座標とともに出力する(1102)。

【0034】類似ウィンドウ抽出部3は、画像分割部2からウィンドウ画像データを入力すると、学習画像データベース41の全ての学習ウィンドウ画像データと差、例えば、各画素値の差の二乗の和、を算出して、最も差の小さいものを抽出する。学習画像データベース41には予め、種々の物体の画像が、図5に示すように、入力ウィンドウ画像と同じサイズのウィンドウに区切られ、ウィンドウ番号とウィンドウの中心点の位置座標とともに格納されている。類似ウィンドウ抽出部3は、全ての入力ウィンドウに対してそれぞれ最も類似した学習ウィンドウを学習画像データベース41から抽出すると、学習ウィンドウが属する学習画像ごとにまとめて、図6に示すように、学習ウィンドウ識別番号と、対応する入力ウィンドウの中心座標の対で出力する(1103)。ここで、学習ウィンドウ識別番号は、図5に示すように、属する学習画像番号と学習画像中の位置を示す番号から成り、また、入力ウィンドウの座標と対応付けることにより、各学習ウィンドウは、ベクトルとみなすことができる。同種画像情報データベース44には、学習画像データベース41に格納されている同じ種類の物体の学習画像について、各物体の同じ部分を含む学習ウィンドウ同士を対応付けて、同種画像情報として格納してある。例えば、図5に示す学習画像3と図10に示す学習画像29は、横向のセダンの学習画像であり、学習画像3の学習ウィンドウ3と学習画像29の学習ウィンドウ27は、両方のセダンの後部ウィンドウを含む画像であるので、対応する学習ウィンドウとみなす。同様にして同種画像情報データベース44には、図11に示すように、学習画像3と学習画像29の学習ウィンドウの対応と、同種の学習画像3、29、62・・・の代表画像が学習画像3であるという情報が格納されている。

【0035】同種画像混合部50は、類似ウィンドウ抽出部3から学習ウィンドウ識別番号と入力ウィンドウ座標を入力すると、同種画像情報データベース44を検索して、同種類の学習画像に属する学習ウィンドウを全て、代表的な1つの学習画像に属する学習ウィンドウで置換する(1104-1)。例えば、学習画像29の学習ウィンドウ29-28は、図11の表に基づいて学習ウィンドウ3-4に変換して、対応する入力ウィンドウ座標とともに出力する。基底選択部51は、学習ウィンドウ識

10

20

30

40

50

13

別番号と座標を入力すると、まず、同じ学習画像に属する学習ウィンドウを抽出する(1104-2)。

【0036】次に、抽出した学習ウィンドウから基底とする3つのウィンドウの組み合わせを全て作る(1105)。例えば、図6の画像3に属する基底は、(3-3, 3-4, 3-11), (3-3, 3-4, 3-12), ...となる。

【0037】次に、各基底ウィンドウの組に対して、基底ウィンドウが張る空間での基底以外のウィンドウの座標を求める(1106)。例えば、図6の基底(3-3, 3-4, 3-11)の基底ベクトルは、3-3を原点とすると、3-3, 3-4から成るベクトル(119-40, 86-100)=(79, -14)(=ベクトルe1)と3-3, 3-11から成るベクトル(54-40, 179-100)=(14, 79)(=ベクトルe2)である。よって、基底以外のウィンドウ3-5(198, 72)の3-3を原点としたときの座標(198-40, 72-100)=(158, -28)をベクトルe1とベクトルe2で表すと、(2\*ベクトルe1) + (0\*ベクトルe2)となり、3-5の相対座標は、(2, 0)となる。ここで学習ウィンドウ座標算出部42は、予め、全ての学習画像の全ての基底に対する各学習ウィンドウの学習画像上での相対座標を、各ウィンドウの中心点の座標から算出して、学習ウィンドウ座標データベース43に格納している。例えば、図6における3-3と3-4と3-5と3-11の学習画像上での中心座標は、図5に示したように、それぞれ(200, 40)(280, 40)(360, 40)(200, 120)であるので、3-3, 3-4, 3-11を基底としたときの基底ベクトルは、3-3を原点とすると(80, 0)=1, (0, 80)=2となる。よって、3-5(360, 40)の、3-3を原点としたときの座標は(160, 0)となり、相対座標は、(2\*ベクトルe1) + (0\*ベクトルe2すなわち、(2, 0)となる。

【0038】集計部53は、(1106)で算出した座\*30

$$e_i = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad \dots (2)$$

(ただし、e1は、入力画像の基底ベクトル、

e2は、学習画像上の基底ベクトルを示す。)

【0042】(実施の形態3)図12は本発明の実施の形態3における画像認識装置のブロック構成図を示す。図12において、1は認識したい対象物の画像データを 40 入力する画像入力部、2は画像入力部1で入力した画像を局所ウィンドウに分割して出力する画像分割部、3は画像分割部2で分割した各入力ウィンドウ画像に対して類似する学習ウィンドウをデータベースから抽出して学習ウィンドウが属する画像ごとにまとめて、それらの位置座標を対応する入力ウィンドウの座標と共に出力する類似ウィンドウ抽出部、4は認識したい物体のモデルを予め作成しておく学習手段、41は種々の物体のモデル画像である学習画像を、画像分割部2で作成する局所ウィンドウと同じサイズの局所ウィンドウに分割して学習

14

\* 標(座標A)と同じ基底・同じウィンドウによる座標を学習ウィンドウ座標データベース43から抽出して(座標B)、座標Aが座標Bの近傍にあれば(1107)、その基底の得点として1点加算する(1108)。例えば、上記の、基底(3-3, 3-4, 3-11)・ウィンドウ3-5の場合は、座標Aと座標Bが一致するので、基底(3-3, 3-4, 3-11)の得点に1点加算する。図7に示すように、全ての画像の全ての基底について(1104)から(1108)までの処理が終了したら、その学習画像の全ての学習ウィンドウについて処理が終了したかを判定する(1109)。次に、その学習画像の全ての基底について処理が終了したかを判定する(1110)。次に、全ての学習画像について処理が終了したかを判定する(1111)。

【0039】対象画像決定部6は、学習画像ごとに得点を合計して最も高得点の学習画像の物体を入力画像の対象物体とする(1112)。

【0040】更に、姿勢算出部62は、その学習画像に属する高得点の基底の入力画像上の座標と、学習画像データベース41に格納されている学習画像上の座標から、物体の平行移動量と回転角を求める(1113)。例えば、図7の高得点の基底(3-3, 3-4, 3-11)の場合、3-3の学習画像上の座標(200, 40)(図5)と、入力画像上の座標(40, 100)(図4)から、平行移動量を(-160, 60)とすると、回転角θは、3-4の学習画像上の座標(280, 40)(図5)と入力画像上の座標(119, 86)(図4)から、(数2)を解いて、θ=10度となる。

【0041】

【数2】

ウィンドウとして格納するとともに、物体上の1点の座標を注目点座標として格納している学習画像データベース、5は類似ウィンドウ抽出部3で抽出した学習ウィンドウの、学習画像上での相対位置と、対応する入力ウィンドウの入力画像上での相対位置を比較して、一致するものの数を集計する投票手段、54は類似ウィンドウ抽出部3が出力する同一学習画像に属する学習ウィンドウの中から、2つの学習ウィンドウの組み合わせを全て抽出して、対応する入力ウィンドウ座標と共に出力するウィンドウ選択部、52-1はウィンドウ選択部54から入力した各2学習ウィンドウについて、それが属する学習画像の注目点座標を学習画像データベース41から入力して、2学習ウィンドウの座標と注目点座標との位置



関係と対応する2入力ウィンドウの座標から入力画像中の物体の注目点座標を算出して、学習画像の注目点座標と共に出力する注目点座標算出部、52-2はウィンドウ選択部54から入力した2学習ウィンドウの座標が作るベクトルと、対応する2入力ウィンドウの座標が作るベクトルとのなす角を算出して出力するベクトル角算出部、53は注目点座標算出部52-1から入力した座標と、ベクトル角算出部52-2から入力した座標を組にして同じ値を持つ組の数を画像ごとに集計して出力する集計部、6は投票手段5の集計結果を受けて入力画像中の対象物とその姿勢を決定する対象決定手段、63は集計部53から入力した集計値のうち最大の値を持つものの学習画像中の物体を入力画像の物体であるとして、その組の座標となす角から入力画像中の対象の位置と回転角を決定する対象姿勢決定部である。

【0043】以上のように構成された画像認識装置について、以下その動作を図12のフローチャートを用いて説明する。図4は入力画像の一例、図5は学習画像の一例、図6は類似ウィンドウ抽出部3が出力するデータの一例である。

【0044】認識対象となる画像データが画像入力部1から入力されると(1401)、画像分割部2は図4に示すように、その画像から一定サイズのウィンドウを順次抽出して、ウィンドウ画像データをウィンドウの中心点の座標とともに出力する(1402)。

【0045】類似ウィンドウ抽出部3は、画像分割部2からウィンドウ画像データを入力すると、学習画像データベース41の全ての学習ウィンドウ画像データと差、例えば、各画素値の差の二乗の和、を算出して、最も差の小さいものを抽出する。学習画像データベース41には予め、種々の物体の画像が、図5に示すように、入力ウィンドウ画像と同じサイズのウィンドウに区切られ、ウィンドウ番号とウィンドウの中心点の位置座標とともに格納されている。また同時に、学習画像ごとに物体上\*

$$(x-40)^2 + (y-100)^2 = 28800$$

$$(x-119)^2 + (y-86)^2 = 16000$$

... (3)

【0048】より求めることができる。ベクトル角算出部52-2は、ウィンドウ選択部54から2ウィンドウの組を入力すると、まず、その2ウィンドウの学習画像上の座標を学習画像データベース41から抽出して、ベクトルを作成する。次に、対応する入力画像上の座標から作ったベクトルとのなす角を算出して出力する(14

\*の任意の1点の座標が注目点座標として格納されている。類似ウィンドウ抽出部3は、全ての入力ウィンドウに対してそれぞれ最も類似した学習ウィンドウを学習画像データベース41から抽出すると、学習ウィンドウが属する学習画像ごとにまとめて、図6に示すように、学習ウィンドウ識別番号と、対応する入力ウィンドウの中心座標の対で出力する(1403)。ここで、学習ウィンドウ識別番号は、図5に示すように、属する学習画像番号と学習画像中の位置を示す番号から成っている。

【0046】ウィンドウ選択部54は、学習ウィンドウ識別番号と座標を入力すると、まず、同じ学習画像に属する学習ウィンドウを抽出する(1404)。次に、抽出した学習ウィンドウから2つのウィンドウの組み合わせを全て作る(1405)。例えば、図6の画像3からは、(3-3, 3-4), (3-3, 3-5), (3-3, 3-11) ... を抽出する。注目点座標算出部52-1は、ウィンドウ選択部51から入力した各2ウィンドウに対して、2学習ウィンドウの座標と学習画像の注目点座標を学習画像データベース41から抽出して、それらの座標間の関係と、対応する入力ウィンドウの座標から入力画像の注目点座標を求める(1406)。例えば、図7のウィンドウ(3-3, 3-4)について、画像3の注目点座標を(320, 160)とすると、図5より、3-3の学習画像上の座標は(200, 40)、3-4の学習画像上の座標は(280, 40)であるから、3-3と注目点の距離の二乗は $(320-200)^2 + (160-40)^2 = 28800$ 、3-4と注目点の距離の二乗は $(320-280)^2 + (160-40)^2 = 16000$ 、となる。図6より、3-3と対応する入力ウィンドウの入力画像上の座標は(40, 100)、3-4と対応する入力ウィンドウの入力画像上の座標は(119, 86)であるから、以上より、入力画像の注目点をx, yとすると、x, yは、

【0047】

【数3】

07)。例えば、図7の(3-3, 3-4)の組の場合、学習画像上の座標(200, 40)(280, 40)(図5)からベクトル(80, 0)入力画像上の座標(40, 100)(119, 86)(図4)からベクトル(79, -14)を得て、なす角

【0049】

【数4】

e1とe1の内積

$$\begin{aligned}\cos \theta &= \frac{\text{(ベクトル } e_1 \text{ の大きさ)} * \text{(ベクトル } e_1 \text{ の大きさ)}}{80 * \sqrt{79^2 + 14^2}} \\ &= \frac{80 * 79}{80 * \sqrt{79^2 + 14^2}} \quad \dots (4)\end{aligned}$$

(ただし、e1は、入力画像の基底ベクトル、

e1は、学習画像上の基底ベクトルを示す。)

【0050】より、 $\theta=10$ 度を得る。集計部53は、注目点座標算出部52-1から入力した座標と、ベクトル角算出部52-2から入力した角度を1つの組にして、同じ値を持つ組の数を画像ごとに集計する(1408)。

【0051】全ての画像の全てのウィンドウについて(1404)から(1408)までの処理が終了したら(1409、1411)、対象姿勢決定部63は、最も高得点の学習画像の物体を入力画像の対象物体とし、高得点となった値の組の座標と角度を入力画像中の物体の位置と回転角とする(1412、1413)。

【0052】(実施の形態4)図14は、本発明の実施の形態4における画像認識装置のブロック構成図を示す。図14において、1は認識したい対象物の画像データを入力する画像入力部、2は画像入力部1で入力した画像を局所ウィンドウに分割して出力する画像分割部、3は画像分割部2で分割した各入力ウィンドウ画像に対して類似する学習ウィンドウをデータベースから抽出して学習ウィンドウが属する画像ごとにまとめて、それらの位置座標を対応する入力ウィンドウの座標と共に出力する類似ウィンドウ抽出部、4は認識したい物体のモデルを予め作成しておく学習手段、41は種々の物体のモデル画像である学習画像を、画像分割部2で作成する局所ウィンドウと同じサイズの局所ウィンドウに分割して学習ウィンドウとして格納するとともに、物体上の1点の座標を注目点座標として格納している学習画像データベース、44は学習画像データベース41中の同じ種類の複数の学習画像についてそれらに属する各学習ウィンドウを代表的な一つの学習画像の同じ部分の学習ウィンドウと対応づけて格納している同種画像情報データベース、5は類似ウィンドウ抽出部3で抽出した学習ウィンドウの、学習画像上での相対位置と、対応する入力ウィンドウの入力画像上での相対位置を比較して、一致するものの数を集計する投票手段、50は同種画像情報データベース44に格納されている同種画像情報に基づいて類似ウィンドウ抽出部3が出力する同一学習画像に属する学習ウィンドウを同じ種類の代表的な1学習画像に含まれる学習ウィンドウに変換して出力する同種画像混合部、54は類似ウィンドウ抽出部3が出力する同一学習

画像に属する学習ウィンドウの中から、2つの学習ウィンドウの組み合わせを全て抽出して、対応する入力ウィンドウ座標と共に出力するウィンドウ選択部、52-1はウィンドウ選択部54から入力した各2学習ウィンドウについて、それが属する学習画像の注目点座標を学習画像データベース41から入力して、2学習ウィンドウの座標と注目点座標との位置関係と対応する2入力ウィンドウの座標から入力画像中の物体の注目点座標を算出して、学習画像の注目点座標と共に出力する注目点座標算出部、52-2はウィンドウ選択部54から入力した2学習ウィンドウの座標が作るベクトルと、2入力ウィンドウの座標が作るベクトルとのなす角を算出して出力するベクトル角算出部、53は注目点座標算出部52-1から入力した座標と、ベクトル角算出部52-2から入力した座標を組にして同じ値を持つ組の数を画像ごとに集計して出力する集計部、6は投票手段5の集計結果を受けて入力画像中の対象物とその姿勢を決定する対象決定手段で、63は集計部53から入力した集計値のうち最大の値を持つものの学習画像中の物体を入力画像の物体であるとして、その組の座標となす角から入力画像中の対象の位置と回転角を決定する対象姿勢決定部である。

【0053】以上のように構成された画像認識装置について、以下その動作を図15のフローチャートを用いて説明する。図4は入力画像の一例、図5は学習画像の一例、図7は類似ウィンドウ抽出部3が出力するデータの一例、図10は学習画像データベース41に格納されている同種画像の一例、図11は同種画像情報データベース44に格納されている同種画像情報の一例である。

【0054】認識対象となる画像データが画像入力部1から入力される(1501)。画像分割部2は、図4に示すように、その画像から一定サイズのウィンドウを順次抽出して、ウィンドウ画像データをウィンドウの中心点の座標とともに出力する(1502)。

【0055】類似ウィンドウ抽出部3は、画像分割部2からウィンドウ画像データを入力すると、学習画像データベース41の全ての学習ウィンドウ画像データと差、例えば、各画素値の差の二乗の和、を算出して、最も差の小さいものを抽出する。学習画像データベース41に

は予め、種々の物体の画像が、図 5 に示すように、入力ウィンドウ画像と同じサイズのウィンドウに区切られ、ウィンドウ番号とウィンドウの中心点の位置座標とともに格納されている。また同時に、学習画像ごとに物体上の任意の 1 点の座標が注目点座標として格納されている。類似ウィンドウ抽出部 3 は、全ての入力ウィンドウに対してそれぞれ最も類似した学習ウィンドウを学習画像データベース 4 1 から抽出すると、学習ウィンドウが属する学習画像ごとにまとめて、図 7 に示すように、学習ウィンドウ識別番号と、対応する入力ウィンドウの中心座標の対で出力する (1503)。ここで、学習ウィンドウ識別番号は、図 5 に示すように、属する学習画像番号と学習画像中の位置を示す番号から成っている。同種画像情報データベース 4 4 には、学習画像データベース 4 1 に格納されている同じ種類の物体の学習画像について、各物体の同じ部分を含む学習ウィンドウ同士を対応付けて、同種画像情報として格納してある。例えば、図 5 に示す学習画像 3 と図 1 2 に示す学習画像 29 は、横向のセダンの学習画像であり、学習画像 3 の学習ウィンドウ 3 と学習画像 29 の学習ウィンドウ 27 は、両方のセダンの後部ウィンドウを含む画像であるので、対応する学習ウィンドウとみなす。同様に同種画像情報データベース 4 4 には、図 1 1 に示すように、学習画像 3 と学習画像 29 の学習ウィンドウの対応と、同種の学習画像 3、29、62・・・の代表画像が学習画像 3 であるという情報が格納されている。

【0056】同種画像混合部 50 は、類似ウィンドウ抽出部 3 から学習ウィンドウ識別番号と入力ウィンドウ座標を入力すると、同種画像情報データベース 4 4 を検索して、同種類の学習画像に属する学習ウィンドウを全

\* て、代表的な 1 つの学習画像に属する学習ウィンドウで置換する (1504-1)。例えば、学習画像 29 の学習ウィンドウ 29-28 は、図 1 1 に基づいて学習ウィンドウ 3-4 に変換して、対応する入力ウィンドウ座標とともに出力する。

【0057】ウィンドウ選択部 54 は、学習ウィンドウ識別番号と座標を入力すると、まず、同じ学習画像に属する学習ウィンドウを抽出する (1504-2)。

【0058】次に、抽出した学習ウィンドウから 2 つのウィンドウの組み合わせを全て作る (1505)。例えば、図 7 の画像 3 からは、(3-3, 3-4), (3-3, 3-5), (3-3, 3-11)・・・を抽出する。

【0059】注目点座標算出部 52-1 は、ウィンドウ選択部 54 から入力した各 2 ウィンドウに対して、2 学習ウィンドウの座標と学習画像の注目点座標を学習画像データベース 4 1 から抽出して、それらの座標間の関係と、対応する入力ウィンドウの座標から入力画像の注目点座標を求める (1506)。例えば、図 6 のウィンドウ (3-3, 3-4) について、画像 3 の注目点座標を (320, 160) とすると、図 5 より、3-3 の学習画像上の座標は (200, 40)、3-4 の学習画像上の座標は (280, 40) であるから、3-3 と注目点の距離の二乗は  $(320-200)^2 + (160-40)^2 = 28800$ 、3-4 と注目点の距離の二乗は  $(320-280)^2 + (160-40)^2 = 16000$ 、となる。図 6 より、3-3 と対応する入力ウィンドウの入力画像上の座標は (40, 100)、3-4 と対応する入力ウィンドウの入力画像上の座標は (119, 86) であるから、以上より、入力画像の注目点を  $x, y$  とすると、 $x, y$  は、

【0060】

$$\left. \begin{aligned} (x-40)^2 + (y-100)^2 &= 28800 \\ (x-119)^2 + (y-86)^2 &= 16000 \end{aligned} \right\} \dots (5)$$

【0061】より求めることができる。ベクトル角算出部 52-2 は、ウィンドウ選択部 54 から 2 ウィンドウの組を入力すると、まず、その 2 ウィンドウの学習画像上の座標を学習画像データベース 4 1 から抽出して、ベクトルを作成する。次に、対応する入力画像上の座標から作ったベクトルとのなす角を算出して出力する。(150

7)例えば、図 7 の (3-3, 3-4) の組の場合、学習画像上の座標 (200, 40) (280, 40) (図 5) からベクトル (80, 0) 入力画像上の座標 (40, 100) (119, 86) (図 4) からベクトル (79, -14) を得て、なす角

【0062】

【数 6】

e1とe1の内積

$$\begin{aligned} \cos \theta &= \frac{(\text{ベクトル } e_1 \text{ の大きさ}) \times (\text{ベクトル } e_1 \text{ の大きさ})}{80 \times 79} \\ &= \frac{80 \times 79}{80 \times \sqrt{79^2 + 14^2}} \end{aligned} \quad \dots (6)$$

(ただし、e1は、入力画像の基底ベクトル、

e1は、学習画像上の基底ベクトルを示す。)

【0063】より、 $\theta=10$ 度を得る。集計部53は、注目点座標算出部52-1から入力した座標と、ベクトル角算出部52-2から入力した角度を1つの組にして、同じ値を持つ組の数を画像ごとに集計する(1508)。

【0064】全ての画像の全てのウィンドウについて(1504)から(1508)までの処理が終了したら(1509、1511)、対象姿勢決定部63は、最も高得点の学習画像の物体を入力画像の対象物体とし、高得点となった値の組の座標と角度を入力画像中の物体の位置と回転角とする(1512、1513)。

【0065】

【発明の効果】以上のように本発明によれば、入力画像中の物体が、学習画像の物体と比較して平行移動や画像面内回転移動している場合にも、対象を認識でき、その平行移動量や回転角を算出することができる。

【0066】また、学習画像に同種の物体の画像が複数ある場合でも、入力画像中の物体の種類や平行移動量・回転角を推定することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態1における画像認識装置のブロック構成図

【図2】本発明の実施の形態1におけるコンピュータによる画像認識装置のブロック構成図

【図3】本発明の実施の形態1における処理の流れを示すフローチャート

【図4】本発明の実施の形態における入力画像の一例を示す図

【図5】本発明の実施の形態における学習画像データベースが保管している学習画像データの一例を示す図

【図6】本発明の実施の形態における類似ウィンドウ抽出部が出力する入力ウィンドウと学習ウィンドウの対応の一例を示す図

【図7】集計部が出力する集計の一例を示す図

【図8】本発明の実施の形態2における画像認識装置のブロック構成図

【図9】本発明の実施の形態2における処理の流れを示すフローチャート

【図10】本発明の実施の形態における画像データベー

ス中の同種画像の一例を示す図

【図11】本発明の実施の形態における同種画像情報データベースが保管している同種画像情報の一例を示す図

【図12】本発明の実施の形態3における画像認識装置のブロック構成図

【図13】本発明の実施の形態3における処理の流れを示すフローチャート

【図14】本発明の実施の形態4における画像認識装置のブロック構成図

【図15】本発明の実施の形態4における処理の流れを示すフローチャート

【図16】従来の画像認識装置の一例を示すブロック図

【符号の説明】

- 1 画像入力部
- 2 画像分割部
- 3 類似ウィンドウ抽出部
- 4 学習手段
- 5 投票手段
- 6 対象決定手段
- 41 学習画像データベース
- 42 学習ウィンドウ座標計算部
- 43 学習ウィンドウ座標データベース
- 44 同種画像情報データベース
- 50 同種画像混合部
- 51 基底選択部
- 52 入力ウィンドウ座標算出部
- 52-1 注目点座標算出部
- 52-2 ベクトル角算出部
- 53 集計部
- 54 ウィンドウ選択部
- 61 対象画像決定部
- 62 姿勢算出部
- 63 対象姿勢決定部
- 201 コンピュータ
- 202 CPU
- 203 メモリ
- 204 キーボード/ディスプレイ
- 205 FDユニット
- 206~208 I/Fユニット

209 CPUバス

210 カメラ

211 画像データベース

212 学習画像データベース

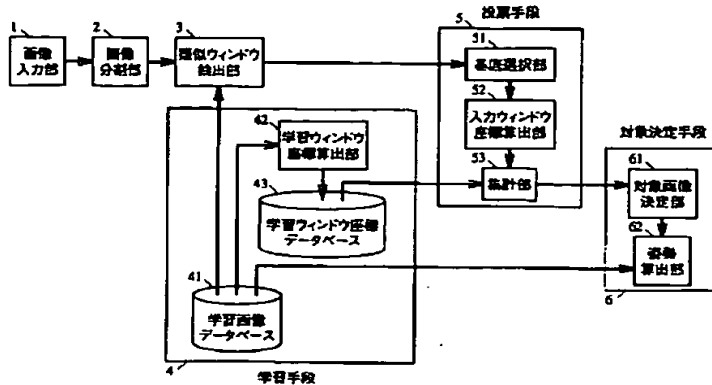
\* 213 学習ウィンドウ座標データベース

214 同種画像情報データベース

215 出力端子

\*

【図1】



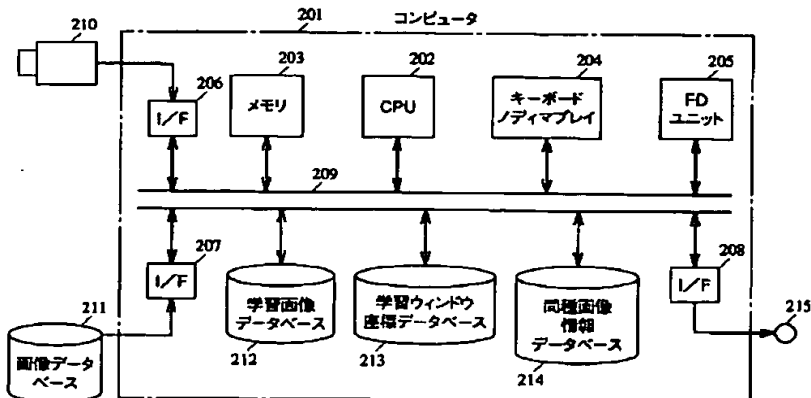
【図6】

学習ウィンドウ座標	入力ウィンドウ座標
3-3	(40, 100)
3-4	(119, 88)
3-5	(188, 72)
.	.
3-11	(54, 179)
3-12	.
63-1	(291, 88)
.	.
72-12	(208, 19)

【図7】

座標	得点
3-3, 3-4, 3-11	19
3-3, 3-4, 3-12	8
.	.
63-1, 63-3, 63-9	1

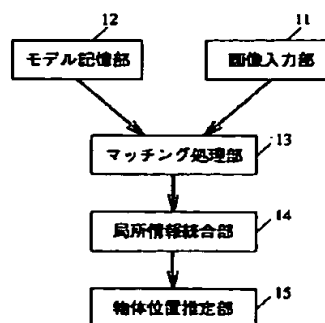
【図2】



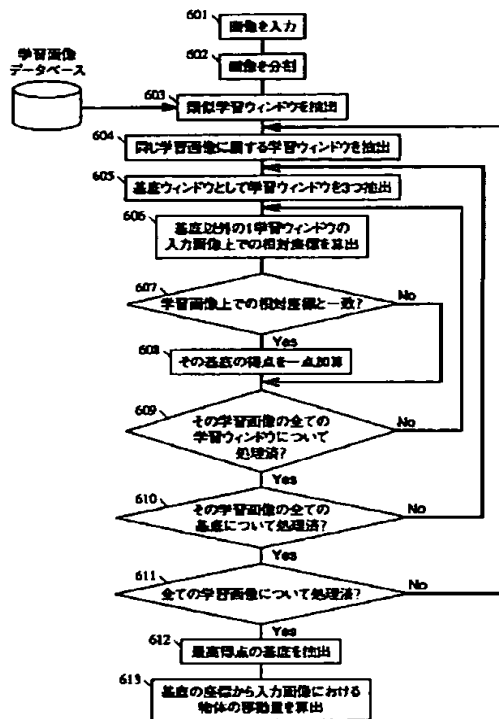
【図11】

代表学習画像の 学習ウィンドウ座標	同種学習画像 画像29	画像62
3-2	29-27	
3-3	29-28	
3-4	29-29	
3-5	29-30	
3-6	29-31	
3-10	29-32	
3-11	29-33	
3-12	29-34	
.	29-35	

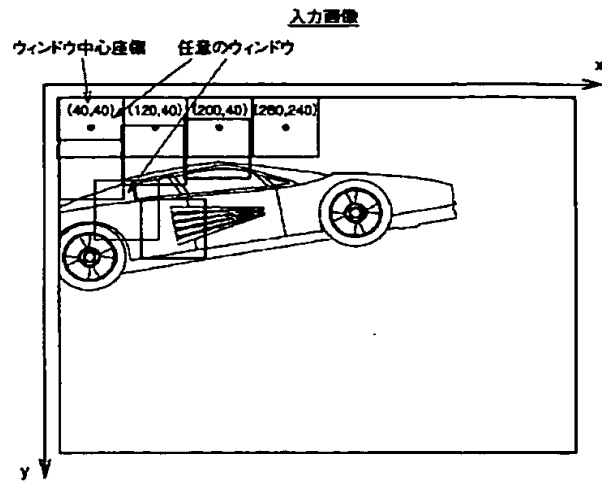
【図16】



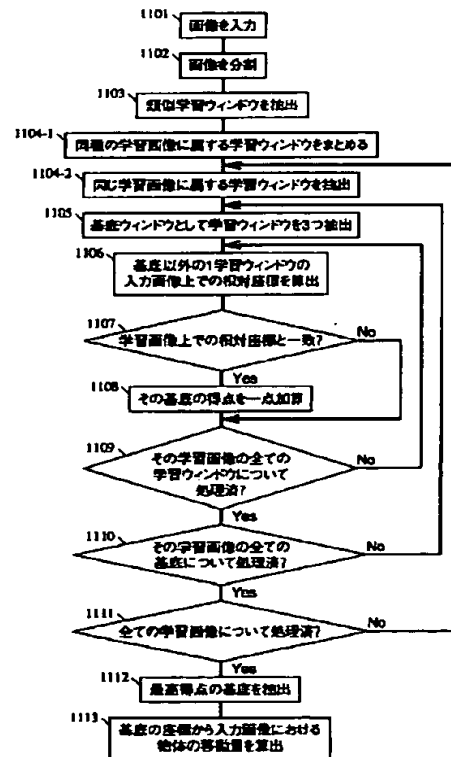
【図3】



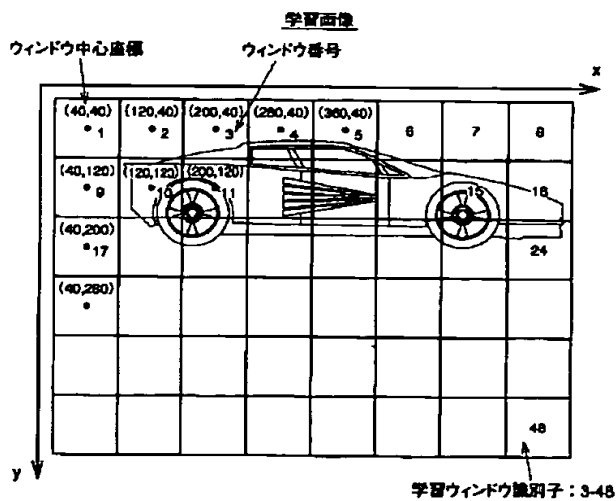
【図4】



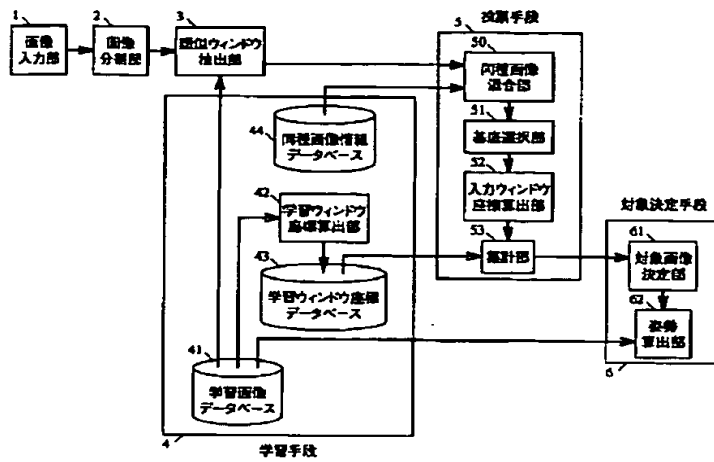
【図9】



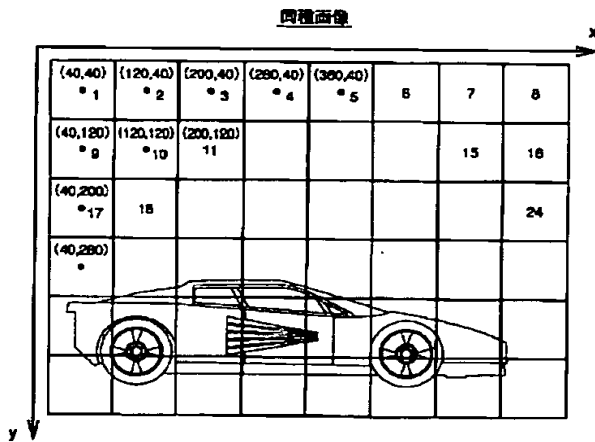
【図5】



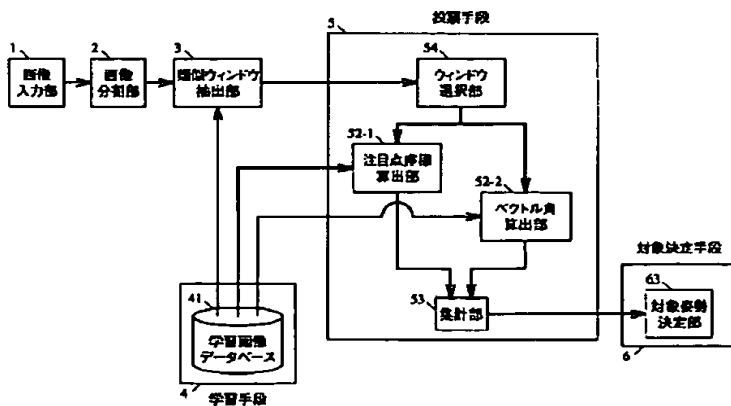
【図 8】



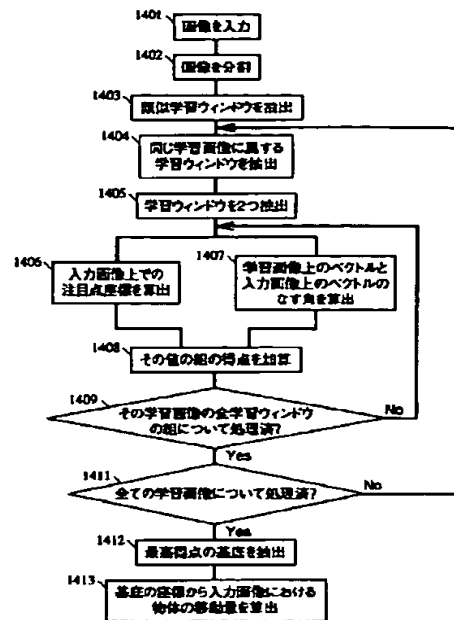
【図 10】



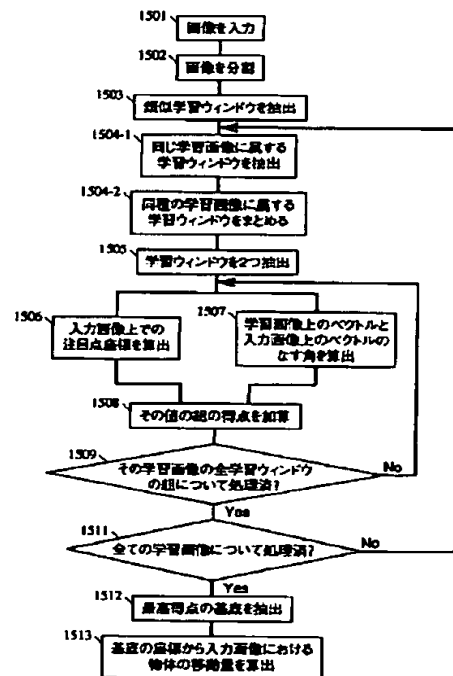
【図 12】



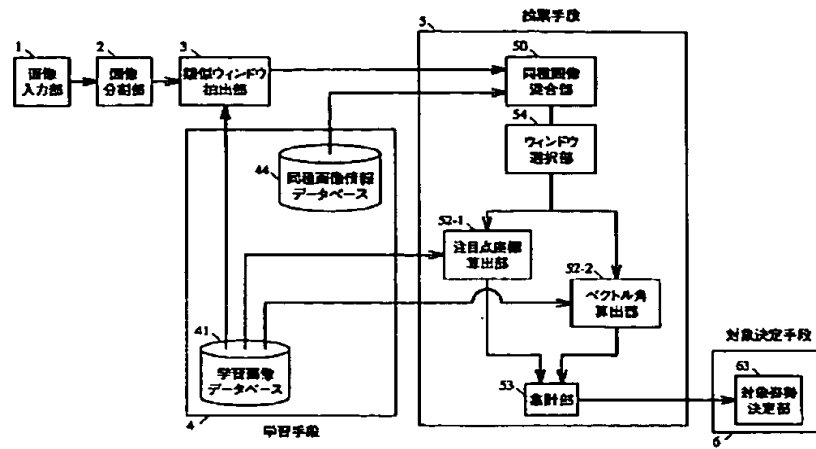
【図 13】



【図 15】



【図14】





**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**